



Análise filogenética: conceitos básicos e suas utilizações como ferramenta para virologia e epidemiologia molecular

Phylogenetic Analysis: Basic Concepts and Its Use as a Tool for Virology and Molecular Epidemiology

Eloiza Teles Caldart^{1,2}, Helena Mata², Cláudio Wageck Canal³ & Ana Paula Ravazzolo²

ABSTRACT

Background: Phylogenetic analyses are an essential part in the exploratory assessment of nucleic acid and amino acid sequences. Particularly in virology, they are able to delineate the evolution and epidemiology of disease etiologic agents and/or the evolutionary path of their hosts. The objective of this review is to help researchers who want to use phylogenetic analyses as a tool in virology and molecular epidemiology studies, presenting the most commonly used methodologies, describing the importance of the different techniques, their peculiar vocabulary and some examples of their use in virology.

Review: This article starts presenting basic concepts of molecular epidemiology and molecular evolution, emphasizing their relevance in the context of viral infectious diseases. It presents a session on the vocabulary relevant to the subject, bringing readers to a minimum level of knowledge needed throughout this literature review. Within its main subject, the text explains what a molecular phylogenetic analysis is, starting from a multiple alignment of nucleotide or amino acid sequences. The different software used to perform multiple alignments may apply different algorithms. To build a phylogeny based on amino acid or nucleotide sequences it is necessary to produce a data matrix based on a model for nucleotide or amino acid replacement, also called evolutionary model. There are a number of evolutionary models available, varying in complexity according to the number of parameters (transition, transversion, GC content, nucleotide position in the codon, among others). Some papers presented herein provide techniques that can be used to choose evolutionary models. After the model is chosen, the next step is to opt for a phylogenetic reconstruction method that best fits the available data and the selected model. Here we present the most common reconstruction methods currently used, describing their principles, advantages and disadvantages. Distance methods, for example, are simpler and faster, however, they do not provide reliable estimations when the sequences are highly divergent. The accuracy of the analysis with probabilistic models (neighbour joining, maximum likelihood and bayesian inference) strongly depends on the adherence of the actual data to the chosen development model. Finally, we also explore topology confidence tests, especially the most used one, the bootstrap. To assist the reader, this review presents figures to explain specific situations discussed in the text and numerous examples of previously published scientific articles in virology that demonstrate the importance of the techniques discussed herein, as well as their judicious use.

Conclusion: The DNA sequence is not only a record of phylogeny and divergence times, but also keeps signs of how the evolutionary process has shaped its history and also the elapsed time in the evolutionary process of the population. Analyses of genomic sequences by molecular phylogeny have demonstrated a broad spectrum of applications. It is important to note that for the different available data and different purposes of phylogenies, reconstruction methods and evolutionary models should be wisely chosen. This review provides theoretical basis for the choice of evolutionary models and phylogenetic reconstruction methods best suited to each situation. In addition, it presents examples of diverse applications of molecular phylogeny in virology.

Keywords: evolution, evolutionary models, molecular epidemiology, phylogeny, phylogenetic reconstruction methods.

Descritores: evolução, modelos evolutivos, epidemiologia molecular, filogenia, métodos de reconstrução filogenética.

I. INTRODUÇÃO

II. CONCEITOS BÁSICOS

1. Epidemiologia Molecular
2. Evolução Molecular e a sua relevância no estudo de Doenças Infecciosas

III. TERMINOLOGIA

IV. ANÁLISE FILOGENÉTICA

1. Alinhamento Múltiplo
 - 1.1. Verificando a saturação nos alinhamentos com sequências de nucleotídeos
2. Modelos Evolutivos
3. Métodos de Reconstrução de Filogenias Moleculares
 - 3.1. Máxima Parcimônia
 - 3.2. Métodos de Distância
 - 3.3. Máxima Verossimilhança
 - 3.4. Inferência Bayesiana
4. Árvores Filogenéticas
5. Teste de Confiança em Topologias

V. IMPORTÂNCIA DA ANÁLISE FILOGENÉTICA COMO FERRAMENTA EM VIROLOGIA E EPIDEMIOLOGIA MOLECULAR

1. Diversidade Molecular
 - 1.1. Coronavírus
 - 1.2. Flavivírus
 - 1.3. Vírus da Raiva
 - 1.4. Lentivírus de Pequenos Ruminantes
 - 1.5. Vírus Influenza A
2. Dinâmica Populacional
 - 2.1. Vírus da Imunodeficiência Felina
3. Evolução Molecular
 - 3.1. Retrovírus
 - 3.2. Vírus Influenza A

VI. CONCLUSÃO

I. INTRODUÇÃO

Há algum tempo, a forma de traçar o caminho percorrido pelas doenças infecciosas eram os exames sorológicos. No entanto, a acumulação acelerada de dados de sequências moleculares de diversos organismos, abre novas possibilidades para o entendimento da evolução das doenças [11]. Análises filogenéticas são, atualmente, uma parte padrão da análise exploratória de sequências de DNA e proteínas, sendo capazes de mostrar o caminho evolutivo e epidemiológico de agentes etiológicos de doenças e/ou o caminho evolutivo de seus hospedeiros [28]. Pesquisadores que dependem de análises comparativas de sequências precisam entender as motivações teóricas e práticas que sustentam novas técnicas e como elas diferem dos métodos precedentes. A capacidade dessas técnicas para tratar de questões anteriormente difíceis está fazendo da análise filogenética uma ferramenta essencial em um número crescente de áreas de pesquisa [42,114]. Além disso, estimar árvores filogenéticas não é uma atividade exclusivamente acadêmica, pois pode ser

útil em investigações forenses [67,71]. O objetivo dessa revisão é auxiliar os pesquisadores que pretendem utilizar a análise filogenética como ferramenta em seus estudos, apresentando as metodologias mais comumente utilizadas, descrevendo a importância das técnicas, seu vocabulário peculiar e alguns exemplos de sua utilização na virologia veterinária.

II. CONCEITOS BÁSICOS

1. Epidemiologia Molecular

A epidemiologia avalia de forma quantitativa a distribuição dos fenômenos de saúde ou doença, seus fatores condicionantes e determinantes nas populações [58]. Permite, ainda, a avaliação da eficácia das intervenções realizadas no âmbito da saúde pública [100]. A epidemiologia em veterinária proporciona a detecção de mudanças nos fatores que interferem na distribuição de determinadas enfermidades, com a finalidade de recomendar e adotar as ações de prevenção e controle adequadas; seja para doenças importantes em produção animal, como para a criação de animais de companhia ou ainda, para doenças com potencial zoonótico [107].

A biologia molecular fornece técnicas altamente sensíveis e específicas, capazes de detectar e analisar a informação presente no genoma dos patógenos e seus hospedeiros. Podemos citar algumas das técnicas de biologia molecular usadas em epidemiologia: *polymerase chain reaction* (PCR), sequenciamento, *restriction fragment length polymorphisms* (RFLP), bancos genômicos, *Random Amplification of Polymorphic DNA* (RAPD), *Polyacrylamide gel electrophoresis* (SDS-PAGE) [19]. A maioria destas técnicas tem por objetivo identificar polimorfismos genéticos gerados, principalmente, pela ocorrência de mutações nos genomas [50].

A definição de epidemiologia molecular compreende o estudo feito sobre indivíduos de determinada população de maneira a se observar sua composição gênica e suas relações com os agravos à saúde; ou seja, a epidemiologia molecular é a ciência que foca na contribuição do potencial genético na etiologia, na distribuição e na prevenção de doenças [83]. Em outras palavras, estuda a evolução molecular de determinado patógeno/gene, inserido em um contexto epidemiológico [33]. Utiliza a hipótese filogenética para inferir como ocorreu a infecção na população: o porquê e de que forma se disseminou [23].

As aplicações da epidemiologia molecular são: identificar os microrganismos responsáveis por doenças infecciosas, determinar suas fontes de infecção e rotas de transmissão, suas relações biológicas (filogenéticas) e identificar os genes (e elementos acessórios) responsáveis por sua virulência, resistência a drogas e indução de imunidade para posterior uso em vacinas [59].

Além da aplicação para o estudo de microrganismos patogênicos, a epidemiologia molecular pode avaliar a relação entre a presença de determinados polimorfismos e a susceptibilidade ou resistência do hospedeiro. O progresso dos estudos nas áreas da biologia molecular e genética está modificando a prática da medicina e da saúde pública através do desenvolvimento de diagnósticos moleculares e de intervenções direcionadas a indivíduos susceptíveis identificados geneticamente [20]. A epidemiologia molecular está diretamente relacionada com a epidemiologia genética e essa proximidade representa a incorporação da bioquímica e da biologia molecular nas pesquisas de epidemiologia tradicional, propondo o acesso a marcadores moleculares [20].

A epidemiologia molecular é uma disciplina emergente que se refere essencialmente à aplicação dos princípios da biologia molecular aos estudos epidemiológicos e, ao mesmo tempo, à utilização das pesquisas e perspectivas epidemiológicas para compreensão das análises moleculares.

2. Evolução Molecular e a Sua Relevância no Estudo de Doenças Infecciosas

Evolução é a mudança de características genéticas em uma população ao longo do tempo. De modo similar a outros sistemas biológicos, parasitas e seus hospedeiros evoluem e o conhecimento sobre processos evolutivos poderá influenciar na maneira de tratar as doenças. Pelo fato das bactérias e vírus se reproduzirem rapidamente, apresentam taxas elevadas de mutações e recombinações gênicas. A seleção natural pode agir rapidamente a favor destas novas variantes, bem como de outras já existentes e mantidas em baixo número na população, como por exemplo, bactérias resistentes a antimicrobianos [6,18].

Um exemplo clássico da colaboração do conhecimento da evolução molecular no tratamento de doenças é o uso de diferentes drogas antirretrovirais associadas (coquetel) no tratamento da AIDs. Esta terapia baseia-se na baixa probabilidade de um vírus

mutante ser resistente a múltiplas drogas ao mesmo tempo [73]. Outro aspecto interessante relacionado à evolução em doenças infecciosas é o entendimento da evolução da virulência [24]. Através do conhecimento da interação entre genes de hospedeiros e seus parasitas é possível identificar fatores importantes para manejo de doenças infecciosas, como taxa de transmissão e virulência. Entretanto, determinar a origem e evolução das doenças infecciosas a partir de sequências gênicas, frequentemente fragmentadas, não é uma tarefa fácil. Por isso, a nossa capacidade de entender a história evolutiva dos organismos depende do entendimento dos métodos pelos quais esta história pode ser inferida [68].

III. TERMINOLOGIA

A fim de abordar as diferentes metodologias da epidemiologia molecular, cabe rever alguns conceitos. O fenótipo dos organismos vivos é o resultado da interação entre a informação genética que eles carregam e passam para a próxima geração com o ambiente em que vivem. O genoma, portador desta informação genética, é, na maioria dos organismos, o ácido desoxirribonucléico (DNA), ao passo que alguns vírus têm o ácido ribonucléico (RNA) como repositório desta informação. Parte da informação genética no DNA é transcrita em RNA: (i) RNA mensageiro (mRNA), que atua como um molde para a síntese de proteínas; (ii) RNA ribossomal (rRNA), que, em conjunto com proteínas ribossomais, constitui o mecanismo de tradução de proteínas; (iii) RNA transportador, que oferece o aminoácido codificado [109]. A história dos organismos é formada por uma série de mudanças evolutivas que têm a peculiaridade de serem hereditárias e, portanto, derivadas da ancestralidade. Diferentes mecanismos de variabilidade genética são os responsáveis pela biodiversidade existente. Esses mecanismos incluem mutações, duplicação de genes, reorganização de genomas e recombinação [108]. De todos esses, principalmente as mutações, deleções e inserções são usadas nos diferentes métodos de filogenia para inferir quais as relações entre os genes. Mutações podem ser a simples troca de uma base nitrogenada por outra, sendo transições mutações entre duas purinas (A e G) ou duas pirimidinas (C e T), enquanto que transversões é a mutação de uma purina por uma pirimidina e vice-versa. Comumente, o número de transições é maior que o de transversões, mas esta proporção pode se alterar em função da

saturação. A saturação pode ser definida como informações históricas (substituições) reescritas várias vezes em sequências de DNA. As deleções (*gaps*) e inserções consistem na retirada de bases nitrogenadas já existentes ou fixação de novas bases ao longo do gene, respectivamente. Essas mutações, também chamadas polimorfismos, podem ser silenciosas ou sinônimas ou não-silenciosas ou não-sinônimas. O primeiro altera a sequência nucleotídica no códon sem alterar o aminoácido formador da proteína; enquanto que o segundo provoca alteração no códon e no aminoácido [50]. O códon é uma sequência de três bases nitrogenadas no mRNA, que especificam determinado aminoácido ou indicam um ponto de parada da tradução. A recombinação é um mecanismo muito importante de variabilidade genética, que consiste na troca de material genético que pode ocorrer durante a meiose ou entre microrganismos presentes no mesmo hospedeiro [110].

O termo homologia é usado quando a característica herdada provém de um ancestral comum. Por outro lado, diversos eventos, como substituições consecutivas, deleções e inserções, podem levar a homoplasia, que consiste em sequências similares sem ancestral comum [36]. Genes originados de um evento de duplicação recente são chamados parálogos. Genes homólogos compartilhados por diferentes espécies que coalescem de um ancestral comum, sem duplicação ou transmissão horizontal, são chamados ortólogos [110].

Uma árvore filogenética (Figura 1) é uma representação gráfica de relações ancestral-descendente entre organismos ou sequências genéticas e deve ser considerada como uma hipótese de um relacionamento evolutivo entre um grupo de organismos. No caso da filogenia molecular, as sequências de nucleotídeos ou proteínas estão nas pontas (*tips*) das árvores, enquanto que os ramos (*branch*) conectam as sequências aos seus ancestrais [42]. Árvores podem ser enraizadas ou não; sendo que, as que possuem raiz contemplam uma sequência chamada grupo externo (*outgroup*), cuja função é dar à árvore uma direção (polaridade) evolutiva, mostrando quem são as sequências mais ancestrais [72]. Em árvores não enraizadas não há nenhuma indicação de qual nó representa o ancestral comum das demais sequências [110]. Os nós de uma árvore filogenética representam um ancestral hipotético e são o ponto de origem dos ramos; os diferentes arranjos dos ramos da

árvore podem dar origem a diferentes grupos (*clusters*). Um dos desafios das inferências filogenéticas é agrupar organismos em grupos monofiléticos que reflitam ancestralidade comum, que difere de homoplasias (por exemplo, por recombinações não-homólogas) e resultantes de diferentes taxas evolutivas.

A distância evolucionária entre duas sequências é o número de mudanças que ocorreram ao longo do ramo da árvore que as representam. Diferentes genes podem ter diferentes taxas evolutivas que, por definição, é o número de substituições de nucleotídeos por sítio, por ano ou milhões de anos [110]. Um sítio corresponde ao local ou posição em que se encontra um determinado nucleotídeo na molécula de DNA [42].

A taxonomia é a parte da biologia voltada para a caracterização da diversidade da vida e organização dos dados que refletem essa diversidade. Táxon é uma unidade taxonômica essencialmente associada a um sistema de classificação, o que significa dizer que pode ser um reino, uma família, um gênero, uma espécie [50]. Ou seja, entidades que de alguma maneira podem ser discernidas umas das outras.

Os marcadores moleculares são sequências de nucleotídeos ou aminoácidos que revelam polimorfismos entre indivíduos geneticamente relacionados e são herdados [32]. Em epidemiologia molecular, são genes ou sequências empregados para rastrear a transmissão de linhagens específicas de agentes etiológicos. Esses dados são normalmente aplicados para descrever a distribuição de cepas em populações e para verificar os fatores de risco de hospedeiros e parasitas na dispersão da doença [36].

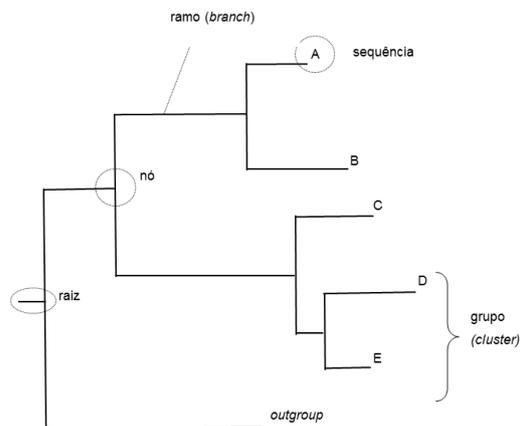


Figura 1. Árvore filogenética enraizada.

IV. ANÁLISE FILOGENÉTICA

Sistemas de classificação de organismos têm sido organizados desde a antiguidade e, desde essa época, já se sabe que é necessário seguir critérios que não variam de pesquisador a pesquisador, evitando a geração de informações conflitantes. E como saber qual o melhor critério a escolher? Após a proposição da teoria da evolução por Darwin, estabeleceu-se como critério a ser escolhido as relações de ancestralidade entre espécies [72]. Sistemas de classificação que levam em conta a evolução entre as espécies podem ser representados através de reconstruções filogenéticas [61,66] e a qualidade dessa estimativa vai depender da qualidade da amostragem e das sequências amostradas. Além da classificação de organismos, análises filogenéticas têm diversas aplicações e entre elas estão: (1) detecção de ortologia ou paralogia, (2) estimação de tempo de divergência, (3) reconstrução de proteínas fósseis, (4) procura por polimorfismos importantes na seleção natural, (5) detecção de pontos de recombinação, (6) identificação de mutações relacionadas a distúrbios ou doenças, (7) determinação da identidade de novos patógenos [42] e (8) determinação da estrutura genética de populações de patógenos inferindo o que produz esta estrutura [43]. Obviamente, para que se possa estabelecer qualquer uma das características acima, devemos utilizar sequências de nucleotídeos e/ou aminoácidos.

1. Alinhamento Múltiplo

A análise filogenética estima a relação entre os genes ou fragmentos de genes inferindo sua história comum. Para fazer isto, é essencial que os locais homólogos sejam comparados uns com os outros e que as sequências utilizadas sejam de boa qualidade [25,26,85]. Portanto, após a obtenção das sequências a serem utilizadas em uma reconstrução filogenética, o alinhamento das mesmas é o primeiro passo para a organização dos dados. Um alinhamento de múltiplas sequências busca homologia posicional entre bases ou aminoácidos [94]. Nós inferimos homologia (ancestralidade comum) quando duas sequências apresentam maior similaridade do que poderia ser esperada se as diferenças tivessem ocorrido ao acaso, o que sugeriria uma origem independente. A representação típica de um alinhamento é uma matriz com as sequências dispostas nas linhas e os sítios nas colunas. Lacunas (*gaps*) são inseridas nas sequências de modo que sítios idên-

ticos ou similares fiquem alinhados na mesma coluna [36]. Minimizando-se, assim, a diferença entre elas. Nem todas as substituições ocorridas, entretanto, serão evidentes no alinhamento. Múltiplas substituições no mesmo sítio de uma sequência mascaram estados intermediários pelos quais a sequência passou. Essas questões tornam a reconstrução da árvore verdadeira necessariamente uma atividade de inferência [42].

Comumente a maior parte dos algoritmos empregados por programas (*softwares*) que realizam alinhamentos múltiplos começam por comparação da similaridade de sequência de todos os pares de sequências, alinhando em primeiro lugar as duas sequências com a maior similaridade. As outras são adicionadas progressivamente. O alinhamento continua acrescentando *gaps* onde forem necessárias para atingir homologia posicional [110]. Quando muitos *gaps* são adicionados aos conjuntos de sequências, o alinhamento total pode ser melhorado por edição manual. A obtenção de um bom alinhamento é um dos passos mais importantes na direção de uma hipótese filogenética confiável [72]. Lacunas no início e no fim do alinhamento ocorrem quando comparamos sequências de diferentes tamanhos [110]. Quando trabalhamos com dados moleculares, frequentemente nos perguntamos se é melhor utilizar sequências de nucleotídeos ou proteínas para obtenção de alinhamento mais acurado. Como regra geral para construção filogenética, se considera uma identidade mínima entre sequências os valores de 70% para nucleotídeos e maior que 25% para aminoácidos, cujos alinhamentos devem conter pelo menos 100 aminoácidos comparados. Se o alinhamento da proteína está com identidade abaixo de 25%, considera-se como um alinhamento não confiável para análise [89]. Da mesma forma, se as sequências de nucleotídeos em um alinhamento apresentarem identidade menor que 70%, sendo uma sequência codificante, é recomendável utilizar a sequência de aminoácidos correspondentes na análise.

1.1. Verificando a saturação nos alinhamentos com sequências de nucleotídeos

Para uma reconstrução filogenética acurada, as sequências devem conter variações, mas estas não podem ser tão divergentes a ponto de acumular substituições múltiplas (saturação) e, conseqüentemente, deixar de refletir um relacionamento filogenético correto [116]. Uma representação visual do grau de saturação em um alinhamento pode ser obtida através

da plotagem do número de transições e transversões observadas *versus* a distância genética corrigida por um modelo de evolução. Simplificando, quando um gráfico representando as distribuições de ambas transições e das transversões apresentam o padrão de uma reta, considera-se que não há saturação das substituições.

2. Modelos Evolutivos

Para construção de uma filogenia a partir de uma sequência de aminoácidos ou nucleotídeos é necessário construir uma matriz de dados baseada em um modelo de substituição de nucleotídeos ou aminoácidos, também chamados modelos evolutivos [78]. Há uma série de modelos evolutivos disponíveis a serem considerados, dos mais simples aos mais complexos. Essa progressão na complexidade é caracterizada pelo aumento do número de parâmetros considerados nas análises.

As estratégias de seleção do modelo evolutivo buscam encontrar o nível adequado de complexidade baseados nos dados disponíveis. [38,70,79]. Certa cautela é necessária quando estamos procurando o melhor modelo para dados heterogêneos. Por exemplo, análises que incluam mais de um gene ou regiões codificantes e não-codificantes: como diferentes regiões genômicas estão sujeitas a diferentes pressões seletivas e limitações evolutivas, um único modelo evolutivo não representará todos os dados. Uma solução é realizar a análise dos dados separadamente para cada gene ou região [77]. O melhor modelo para um conjunto de dados específico pode ser selecionado através de testes estatísticos. O programa JMODELTEST [77] é comumente utilizado para alinhamentos com DNA e o programa ProTest [2], para proteína. Outro ponto importante na escolha do melhor modelo é priorizar um balanço entre acurácia e simplicidade [2]. Portanto, os modelos mais simples devem ser, sempre que possível, preferidos aos mais complexos, pois o aumento no número de parâmetros implica no aumento da variância dos cálculos associados ao modelo [64].

O modelo de substituição “distância p” é o mais simples e representa a proporção de posições em que as duas sequências diferem. Apresenta a menor variância dentre os modelos e deve ser usado quando a taxa evolutiva entre as sequências for constante [64]. Já o modelo Jukes Cantor [49] assume que as substituições são múltiplas e ocorrem aleatoriamente entre os quatro tipos de nucleotídeos, e que os eventos de substituição obedecem à distribuição de probabilidades de *Poisson*. Este modelo assume igualdade

nas probabilidades de substituição entre os diferentes nucleotídeos. No entanto, observa-se que transições são mais frequentes que transversões. Isto pode ser atribuído à considerável diferença no tamanho entre pirimidinas (um anel aromático) e purinas (dois anéis aromáticos), que pode impor restrições de caráter espacial às transversões [36]. O modelo Kimura 2-parâmetros [54] permite levar em conta as substituições múltiplas e, justamente essa desigualdade de ocorrência entre transições e transversões, através do uso de uma taxa P de substituições transicionais e uma taxa Q de substituições transversionais. O modelo Tajima-Nei [103] considera a frequência de C e G nas sequências do alinhamento. Este modelo não distingue transições e transversões. O modelo Tamura 3-parâmetros [105] é uma proposta de extensão ao modelo Kimura 2-parâmetros, para contemplar a observação de que muitas sequências contêm um viés muito forte na composição GC de suas sequências. As terceiras posições (terceiro nucleotídeo) dos códons são particularmente sensíveis a este viés, dado que podem abrigar muitas substituições silenciosas. O modelo HKY [40] é um método probabilístico que contempla a mencionada taxa entre transições e transversões, bem como leva em consideração a frequência de ocorrência de cada uma das bases nas sequências. Tamura-Nei [106] por sua vez, é um modelo que leva em conta duas taxas de transição: transição entre purinas e transição entre pirimidinas. Além de levar em conta a taxa de transversões e a desigualdade na frequência de bases das sequências. Tamura e Nei implementaram este modelo após observar que a taxa de transição entre pirimidinas ($C \leftrightarrow T$) é muito diferente da taxa de transição entre purinas ($G \leftrightarrow A$) na região de controle do DNA mitocondrial de hominídeos, e que a negligência na observação dessa diferença acarretaria em uma subestimativa da distância real [36]. O modelo GTR [116] é o mais complexo (Figura 2). Admite seis diferentes parâmetros, que são as seis possibilidades de substituição entre as bases ($A \leftrightarrow G$; $A \leftrightarrow C$; $A \leftrightarrow T$; $G \leftrightarrow C$; $G \leftrightarrow T$; $C \leftrightarrow T$), além de levar em conta as desigualdades na composição de nucleotídeos entre as sequências.

Os modelos mencionados até então partem da premissa de que a taxa de substituição é constante ao longo da sequência. Contudo, a realidade pode ser bastante distante disso, pois há sítios e regiões extremamente conservados no genoma, indicando que nucleotídeos codificam partes cruciais ao funcionamento de proteínas e também em regiões não traduzidas. Por

outro lado, outros trechos do genoma estarão menos sujeitos a uma pressão de seleção purificadora, eventualmente por não participarem de nenhum sítio funcional ou estrutural crítico protéico ou de regiões não traduzidas. Estes sítios estarão, portanto, mais sujeitos à fixação de mutações [36].

Em geral, os métodos filogenéticos podem ser menos acurados ou inconsistentes quando o modelo evolutivo assumido estiver errado. Se o modelo pode influenciar nos resultados das análises, é crucial decidir com cautela qual o modelo mais apropriado com o qual trabalhar.

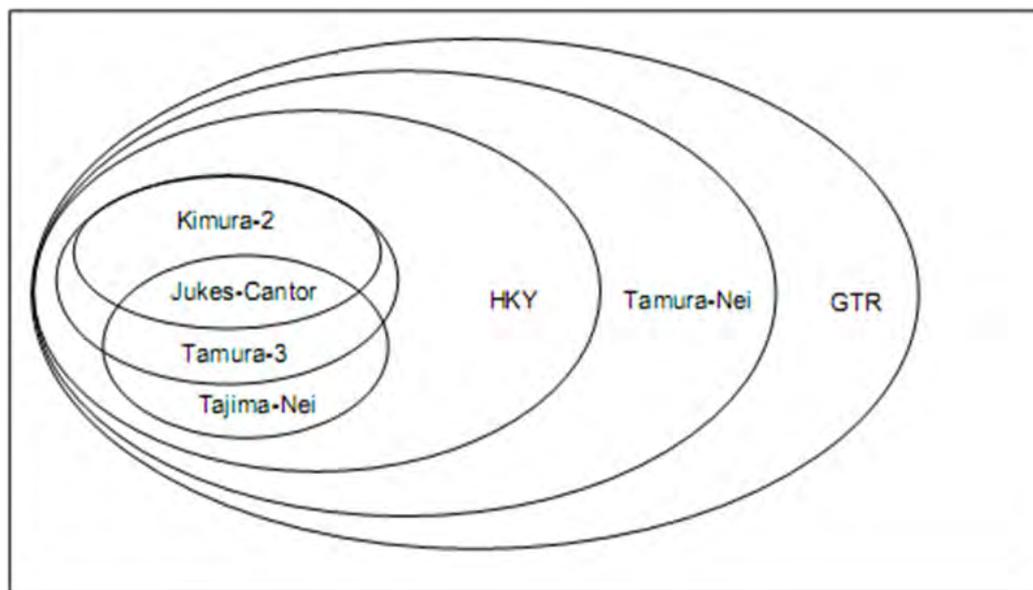


Figura 2. Relações entre os diversos modelos evolutivos, adaptado de Garcia, 2007 [36].

3. Métodos de Reconstrução de Filogenias Moleculares

Os métodos utilizados em filogenia molecular são baseados em suposições de como o processo de evolução ocorreu [31]. Essas suposições podem estar implícitas, como no método de Máxima Parcimônia (MP) (não trabalha com modelos evolutivos), ou explícitas, como nos Métodos de Distância e Máxima Verossimilhança (ML, para *Maximum Likelihood*). Eles descrevem as diferentes possibilidades de mudança de um nucleotídeo ou amino ácido para o outro, com o objetivo de corrigir alterações invisíveis ao longo da filogenia [114]. Talvez o aspecto mais frustrante da análise filogenética para os iniciantes seja a variedade de métodos de inferência que podem ser utilizados e que estão ativamente propostos por diferentes especialistas [46].

Após a escolha do modelo evolutivo, discutidos no item anterior, o próximo passo é optar por um método de reconstrução filogenética que melhor se adeque aos dados disponíveis e ao modelo escolhido. Os métodos conhecidos para a construção de filogenias

são: baseados em matrizes de distâncias (a matriz de dados, nucleotídeos, por exemplo, é transformada em uma matriz de distância) ou baseados em análise de estados do caractere (nucleotídeo ou outro caractere são analisados diretamente, ou seja, sítio a sítio) [94]. O método mais conhecido baseado em distância denomina-se *Neighbor-joining* (agrupamentos vizinhos, NJ). Já entre os denominados métodos discretos, ou probabilísticos estão os métodos de Máxima Parcimônia (MP), Máxima Verossimilhança (ML – *Maximum Likelihood*) e Inferência Bayesiana (IB) [42]. Apesar da existência de diferentes métodos de construção de uma árvore filogenética, existem alguns princípios importantes e comuns aos métodos citados: 1) supõe que características similares são homólogas; 2) baseia-se em caracteres derivados (apomórficos), ou seja, características exclusivas a um grupo de espécies e presente no ancestral comum mais recente deste grupo; 3) relaciona características derivadas compartilhadas como marcadores de parentesco [33]. É importante lembrar que a acurácia das análises com modelos

probabilísticos (NJ, ML e IB) depende fortemente da aderência dos dados reais ao modelo de evolução escolhido [104].

3.1. Máxima Parcimônia

Esse método é definido como o método que considera que uma mutação é mais provável do que duas. É um método discreto, pois atua diretamente nas sequências ou nas funções operadas por elas, mas não faz uso de modelos probabilísticos de evolução. O método de máxima parcimônia busca escolher a árvore, dados os táxons da análise, que explique a evolução entre as sequências com o menor número possível de

substituições [36]. Essa abordagem tem um forte apelo pela simplicidade do raciocínio, inspirado na “navalha de Occam”, princípio que diz que os fenômenos naturais devem ser explicados da maneira mais simples possível, com o menor número de hipóteses e parâmetros [90]. Se o número de mudanças por sítio for relativamente baixa, a MP se aproxima da ML e a topologia das árvores estimadas serão similares [101]. Um potencial viés existe quando a árvore verdadeira tiver ramos internos curtos e ramos terminais longos (Figura 3a), podendo ocorrer uma interpretação incorreta (Figura 3b), em que os ramos longos parecem mais proximamente relacionados do que realmente são [114].

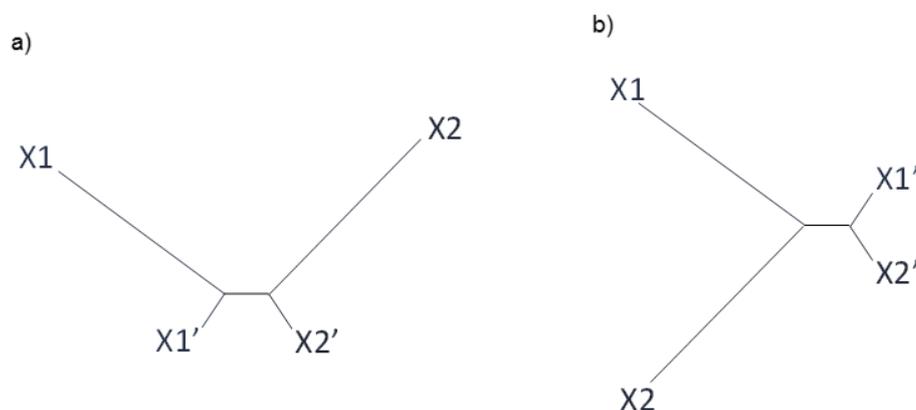


Figura 3. Exemplo clássico de uma topologia hipotética de quatro táxons contendo dois ramos periféricos longos e demais ramos curtos. A árvore verdadeira (Figura 3a) pode sofrer uma interpretação incorreta (Figura 3b), em que os ramos longos parecem mais proximamente relacionados do que realmente são.

3.2. Métodos de Distância

Métodos de distância reduzem a variação entre duas sequências a uma única medida de distância entre elas e trabalham com essas distâncias na estimativa da árvore final. Eles diferem entre si na forma como utilizam a matriz de distâncias para inferir a filogenia. Os métodos de quadrados mínimos e evolução mínima estabelecem critérios relacionados a estas distâncias [14] e buscam a árvore que maximiza o critério estabelecido. Os métodos de agrupamento de distâncias, como UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*) e *Neighbor-Joining* (NJ), utilizam abordagens mais diretas. Em vez de buscar a árvore, dentre todas as possíveis, que otimiza um critério global estabelecido, constroem a árvore nó a nó escolhendo a cada passo um par de nós que serão conectados, sem voltar atrás nessa escolha.

O algoritmo NJ [92,102] é extremamente popular, rápido e apresenta boa performance quando a divergência entre as sequências é baixa. Esse algoritmo é baseado em uma matriz de distância e faz uso de modelos probabilísticos de evolução [42]. A análise pode ser iniciada com todos os táxons ligados a um único nó central configurando o que chamamos de árvore-estrela (não enraizada). A cada passo é escolhido um par de nós para ser agrupado num novo nó, até que toda a árvore tenha sido resolvida. A escolha do par a ser agrupado é baseada no seguinte critério: para cada possível par de nós ainda ligados ao nó central, calcula-se qual par que, agrupado, determinaria a menor soma de comprimentos de ramo da árvore. Após escolhido o par que determina a menor soma total de comprimento de ramos, rearranja-se a matriz de distâncias. Todo o processo é repetido para os nós restantes, até que

a árvore esteja completamente resolvida. Teremos então uma árvore não-enraizada, com comprimentos assinalados em cada ramo [36]. Em suma, a árvore, neste método, é estimada através do comprimento dos ramos com base nas distâncias evolutivas. O método não examina todas as topologias possíveis, mas procura encontrar vizinhos que minimizem o comprimento total da árvore. Simulações mostram que o método de NJ é bastante rápido e eficiente, mesmo quando comparado a métodos substancialmente mais complexos [53,91,104].

3.3. Máxima Verossimilhança

A reconstrução acurada das relações entre organismos que foram separados por um longo tempo, ou estão a evoluir rapidamente, requer um método que corrija vários eventos mutacionais no mesmo sítio [42]. A Máxima Verossimilhança (ML – *Maximum Likelihood*) é um método discreto, assim como a Máxima Parcimônia; no entanto, busca dentre as árvores a mais verossímil de acordo com os dados fornecidos (sequências do alinhamento) e com base no modelo evolutivo. Esse método permite calcular estimativas de menor variância, conseqüentemente, com menos erros de amostragem [29]. A verossimilhança de uma topologia, dado um alinhamento, é a soma das verossimilhanças dessa topologia para cada um dos sítios do alinhamento. A verossimilhança para cada um dos sítios deve ser calculada como o produto da soma das probabilidades de ocorrência de todas as substituições possíveis em um nó. Para todos os nós da árvore serão consideradas todas as possibilidades de comprimento de ramo [21,29].

A grande atração dessa abordagem da probabilidade em filogenia é a existência de uma ferramenta estatística poderosa. Além disso, essa metodologia permite a inferência de árvores filogenéticas usando modelos evolucionários complexos – incluindo a capacidade de estimar os parâmetros do modelo e assim fazer inferências simultaneamente com relação aos padrões e processos da evolução, além de fornecer os meios para a comparação de árvores concorrentes e modelos [114]. A ML é um método muito atrativo, mas também tem desvantagens, pois exige maior esforço computacional e na grande maioria das análises o espaço/tempo necessário para gerar todas as possíveis árvores é imensamente grande e proibitivo para uma busca exaustiva, demandando a adoção de algoritmos heurísticos de otimização de busca.

3.4. Inferência Bayesiana

A inferência Bayesiana é um método probabilístico desenvolvido por Thomas Bayes no século XVIII [5]. A análise Bayesiana é baseada na probabilidade *a posteriori*, utilizando uma probabilidade *a priori* e gerando uma árvore filogenética de acordo com os dados. Na maior parte dos casos, todas as árvores são consideradas, *a priori*, igualmente prováveis e a probabilidade é calculada através de uma série de modelos padrão de evolução de nucleotídeos ou aminoácidos [46]. A probabilidade prévia pode ser usada para incorporar informações básicas sobre a topologia da árvore [80,87] ou sobre o processo de substituição de nucleotídeos, o que significa que a inferência bayesiana permite mudanças à medida que novos dados ou informações são adicionados à análise, mas a maioria dos investigadores opta por não adicionar essas informações.

Entretanto, a busca pela árvore mais provável, levando em consideração todos os tamanhos de ramos, todos os valores de parâmetros para modelos evolutivos e suas combinações, não seria possível. A solução para esta limitação é a estimativa da probabilidade posterior por aproximação através de amostragem por Cadeias Markovianas (*Markov chain Monte Carlo* - MCMC) [37,86]. A ideia central é de fazer pequenas mudanças aleatórias nos parâmetros e depois aceitá-las ou rejeitá-las, de acordo com a probabilidade [88].

4. Árvores Filogenéticas

O objetivo de filogenias moleculares é converter informações inicialmente dadas em sequências de DNA ou aminoácidos em uma árvore evolucionária. Árvores filogenéticas são hipóteses de relações evolucionárias que podem ser representadas de diversas formas (Figura 4): cladogramas, os quais demonstram qual o ancestral mais próximo; filogramas, que mostram o número de substituições que ocorreram por meio de um coeficiente de similaridade; e ultramétricas, que podem ser usadas para representar o tempo evolucionário propriamente dito [72].

5. Teste de Confiança em Topologias

A principal fraqueza de grande parte dos métodos de reconstrução filogenética discutidos até agora é que eles produzem apenas estimativas pontuais da filogenia; um programa de computador é executado e o resultado é uma árvore. A questão imediata é: com que força os dados apoiam cada uma das relações representadas na árvore? Essa questão é comumen-

te respondida por uma técnica estatística chamada *bootstrapping* [22]. De acordo com Felsenstein (1985) [30], o *bootstrap* é o teste mais usado para se avaliar a confiança numa filogenia molecular. Ele permite avaliar a confiança no suporte de cada nó da topologia escolhida, mediante repetição da análise filogenética sobre pseudo-réplicas do alinhamento original. As pseudo-réplicas têm o mesmo tamanho do alinhamento original e suas colunas são sorteadas dentre as colunas do alinhamento original, com possibilidade de repetição, ou seja, uma coluna do alinhamento original pode ser sorteada mais de uma vez. O número de sorteios é igual ao número total de sítios. Desta forma, alguns sítios serão amostrados mais de uma vez enquanto outros estarão fora do pseudo-alinhamento. Portanto, árvores-réplicas são construídas, cada uma baseada

em conjunto de dados diferentes. No final, o teste compara cada uma das árvores-réplicas com a árvore original e o valor do *bootstrap* obtido corresponde ao percentual de vezes que o grupamento foi recuperado nas árvores-réplicas [64]. Geralmente, esse valor é dado em porcentagens sobre os respectivos nós das árvores filogenéticas. A interpretação exata do limite aceitável da proporção de *bootstrap* é abstrata; quanto maior a proporção é fácil interpretar que é maior a confiança na topologia daquele nó. Alguns trabalhos [41,118] concluíram que um ponto de corte razoável seria de 70%. Às vezes, o *bootstrap* pode dar suporte substancial para várias árvores que não são topologicamente similares. Em tais casos, apresentar mais de uma árvore pode ser a única maneira de resumir de forma adequada os dados [42].

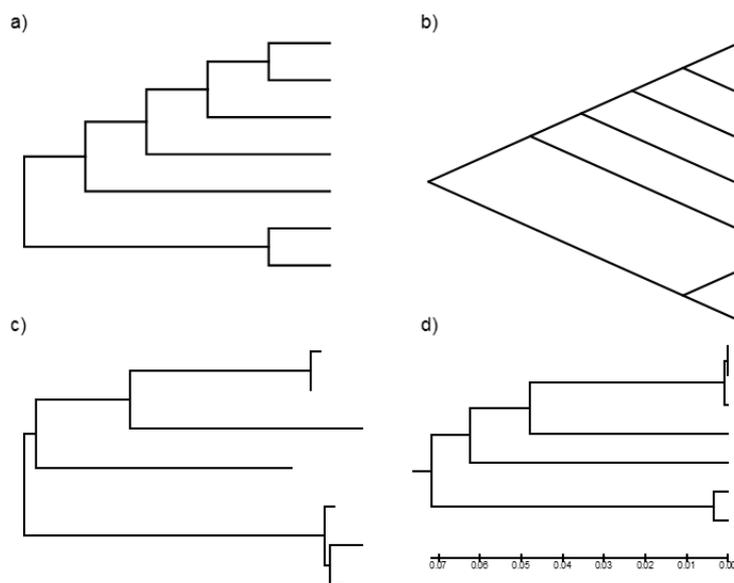


Figura 4. Formas de representação das árvores filogenéticas. 4a- cladograma retangular construído com método NJ; 4b- cladograma clássico construído com método NJ; 4c- filograma representando relacionamento filogenético construído com método NJ e 4d- árvore construída pelo método de distância UPGMA representando uma árvore ultra métrica.

V. IMPORTÂNCIA DA ANÁLISE FILOGENÉTICA COMO FERRAMENTA EM VIROLOGIA E EPIDEMIOLOGIA MOLECULAR

A evolução em duas áreas, principalmente, permitiu grandes avanços na epidemiologia molecular. Em primeiro lugar, os modelos matemáticos de evolução de sequências foram muito melhorados com a incorporação do conhecimento biológico, bioquímico e evolucionário. Em segundo lugar, os testes estatísticos tornaram-se ferramentas indispensáveis para filogeneti-

cistas, permitindo uma avaliação rigorosa das hipóteses evolutivas complexas [114]. Análises filogenéticas podem ser usadas com diversos objetivos, como por exemplo, detecção de ortologia ou paralogia e estimação de tempo de divergência em estudos evolutivos; identificação de polimorfismos importantes na seleção natural, que podem aumentar a imunogenicidade de determinadas proteínas auxiliando na produção de vacinas; detecção de pontos de recombinação para funcionarem como marcadores moleculares na de-

teção de novas linhagens; identificação de mutações relacionadas a distúrbios ou doenças; determinação da identidade de novos patógenos e caracterização molecular [42]. Diferentes abordagens são utilizadas em virologia e algumas delas serão descritas a seguir.

1. Diversidade Molecular

Um grande número de trabalhos científicos tem se dedicado a caracterizar, de forma molecular, os vírus que parasitam diversos reinos com a intenção de estimar a diversidade molecular existente. Adicionalmente, muitos desses trabalhos trazem diversos outros tipos de informação além da caracterização molecular. Alguns exemplos veremos a seguir.

1.1. Coronavírus

Coronavírus são vírus cujo material genético é RNA, sendo conhecidos por infectar várias espécies de mamíferos e aves. Esses vírus possuem um histórico de adaptação a novos hospedeiros, por isso o estudo genético das diferentes espécies e variantes virais contribui para um melhor entendimento das características virais bem como monitorar sua evolução.

A análise filogenética de sequências de Coronavírus bovino (BCoV) de propriedades do sudeste brasileiro, juntamente com sequências de outros países, revelou a formação de dois *clusters*. O *cluster* I incluía as amostras dos outros países e uma amostra brasileira. No *cluster* II houve segregação das amostras portadoras de uma deleção de 18 nucleotídeos, a qual era somente encontrada em sequências de Coronavírus humano (HCoV-OC43), reforçando a hipótese recentemente lançada de potencial zoonótico de BCoV. Além disso, o fragmento sequenciado nesse estudo, na região hipervariável do gene da glicoproteína espiculada, foi eleito um bom marcador genético a ser usado em maiores estudos sobre eficiência de imunógenos, pois é a proteína mais exposta à pressão seletiva imunológica, portanto, a mais propensa a polimorfismos [8]. A diversidade molecular de Coronavírus em morcegos e cães no Brasil também foi inferida através do uso de análises filogenéticas [9,10,74].

A bronquite Infecciosa das galinhas tem provocado grandes perdas econômicas na avicultura em todo o mundo, apesar do uso difundido de vacinas. Amostras do Vírus da Bronquite Infecciosa das Galinhas (*Coronaviridae*) isolados de conteúdo entérico de aves de estabelecimentos de produção brasileiros foram sequenciados [111]. Mais de 20 sorotipos do vírus são

conhecidos e essa diversidade de sorotipos pode ser responsável pelas falhas vacinais e, conseqüentemente, novas epidemias. No Brasil apenas uma vacina é aceita pelo Ministério da Agricultura, a do sorotipo Massachusetts; no entanto, diversos outros sorotipos são encontrados no país. Por meio da caracterização molecular, os autores comprovaram a circulação de outros sorotipos no país e demonstraram a necessidade de uso de vacinas com outras cepas. A análise filogenética mostrou que as amostras segregaram junto ao sorotipo D274 e distante do sorotipo Massachusetts, demonstrando os prováveis motivos da ineficiência do programa de controle da doença no Brasil.

1.2. Flavivirus

A família *Flaviviridae* consiste em um vírus RNA senso positivo e envelopado, sendo composta por três gêneros: *Pestivirus*, *Hepacivirus* e *Flavivirus*, este último contendo mais de 70 espécies [65].

No exemplo abaixo descrevemos sucintamente um estudo filogenético de amostras de Flavivirus brasileiras [4]. Os fragmentos utilizados compreendiam uma porção do gene NS5 até uma região não codificante 3', a qual possui regiões conservadas e hipervariáveis na mesma proporção, tornando-se uma fonte rica de informações filogenéticas para o problema do estudo: divergência entre Flavivirus amostrados no Brasil. Além das amostras brasileiras, sequências do GenBank foram inseridas nas análises. A maior parte das amostras brasileiras agruparam-se em dois ramos, o primeiro da febre-amarela (YF) e o segundo dividido em dengue subtipos 1, 2 e 4. Houve um ramo que apresentou segregação: JEV (*Japanese Encephalitis Virus Complex*), SLE (*Saint Louis Encephalitis*), Ilhéus, Rocio e Cacipacoré. O vírus Iguape apareceu como um separado e distante ramo, demonstrando a divergência entre esses subtipos virais.

1.3. Vírus da Raiva

A raiva causa milhares de mortes por ano apesar da existência de ferramentas que podem prevenir e controlar essa doença. Embora os cães sejam a fonte da maior parte dos casos de raiva, particularmente na África e na Ásia, morcegos são as principais causas de infecção nas Américas [115]. Conhecer as variantes virais e onde elas estão circulando é importante para traçar medidas preventivas.

Amostras provenientes de casos de raiva transmitida por morcegos hematófagos ou relacionados com

esses animais, foram caracterizadas na Argentina [17]. Identificou-se a circulação de quatro variantes antigênicas. Através da análise filogenética verificou-se que duas variantes circulam independentemente no ciclo urbano. Com os dados, os autores identificaram as linhagens atuantes em cada região do país, inclusive em regiões consideradas livres da doença. Em outro trabalho, o gene da nucleoproteína do vírus da raiva de amostras provenientes de animais domésticos (cães e gatos) e canídeos silvestres do Nordeste do Brasil foi estudado [13]. Geralmente, amostras de cães e gatos pertencem a variante antigênica 2, e foi o que aconteceu nesse trabalho. No entanto, dentre as amostras de canídeos silvestres, três se enquadraram na variação antigênica 2, enquanto que as outras nove apresentaram padrão antigênico desconhecido. Na análise filogenética percebemos claramente a presença de dois distintos ciclos de raiva ocorrendo entre canídeos no Nordeste do Brasil: uma relacionada a canídeos silvestres, outra relacionada aos domésticos.

1.4. *Lentivírus de Pequenos Ruminantes*

Desde a descrição das primeiras sequências completas dos genomas de lentivírus de pequenos ruminantes (SRLVs para *Small Ruminant Lentivirus*), vírus Maedi Visna dos ovinos - MVV [99] e vírus da artrite encefalite caprina - CAEV [93], diversos grupos desenvolveram pesquisas a fim de avaliar a sua diversidade, com diferentes finalidades. Desde a avaliação de variabilidade para fins de diagnóstico [39], até a determinação de uma fonte de infecção [96]. Outro aspecto importante das análises filogenéticas realizadas em SRLVs foi a observação de que isolados obtidos de caprinos e ovinos podiam formar um mesmo grupo, indicando a possibilidade de infecção interespecie em regiões em que as duas espécies fazem parte de rebanhos mistos [51,57]. Embora se observe algumas características de sequências de nucleotídeos que, em regiões onde não há contato entre ovinos e caprinos, possam diferenciar a origem das amostras [75,81]. Os primeiros lentivírus isolados de pequenos ruminantes no Brasil foram analisados filogeneticamente [81] juntamente com outras sequências de SRLVs contidas em bases de dados. O isolado brasileiro ovino agrupou com o protótipo MVV e os isolados caprinos com o protótipo CAEV, pelo fato de os isolados terem sido obtidos na região Sul, onde os rebanhos não são mistos. Em contraste com os vírus caprinos, todas as sequências de ovinos

continham uma deleção de seis nucleotídeos no gene *gag*, o que resultou na deleção de dois aminoácidos da região central da proteína do capsídeo (CA). Essa deleção pode ser um marcador útil na análise de SRLV, especialmente quando se considera a possibilidade de transmissão de lentivírus entre ovinos e caprinos. Resultados similares foram descritos em isolados de caprinos da região Norte [27]. Uma análise filogenética em rebanhos de caprinos leiteiros do norte da Itália foi realizada para servir como base de futuros estudos de epidemiologia molecular, com interesse no monitoramento dos efeitos do avanço da doença e futuros programas de erradicação da mesma [75]. A região do genoma analisada foi um pequeno fragmento dentro do gene *gag*, o qual codifica a proteína matriz (MA). Os resultados obtidos, a partir de 52 novas sequências, demonstraram que as amostras italianas segregaram em grupos muito próximos e que são similares ao protótipo CAEV-CO. Isso sugere uma origem comum ao SRLV encontrado nos rebanhos monitorados. Além disso, as amostras analisadas apresentaram alta variabilidade no fragmento estudado e uma inserção de 7 aminoácidos em todas as amostras isoladas de ovinos (MVV), o que possibilita a diferenciação de amostras isoladas de caprinos (CAEV). Os autores concluíram que o fragmento MA é adequado a estudos filogenéticos e pode ser aplicado em programas de monitoramento e erradicação. Em outro estudo com SRLVs, a co-infecção por MVV e CAEV em duas cabras naturalmente infectadas foi descrita [76], sugerindo, inclusive, a ocorrência de recombinação entre esses dois vírus *in vivo*, em um dos animais.

1.5. *Vírus Influenza A*

Influenza aviária é uma doença infecciosa de aves causada pelo vírus influenza A que normalmente não infecta humanos. No entanto, em algumas circunstâncias cepas altamente patogênicas causam doença respiratória grave em humanos. Existe a preocupação de que o vírus possa sofrer uma mutação e, assim, se tornar mais facilmente transmissível entre humanos, aumentando a possibilidade de uma pandemia de gripe [115].

Um estudo de caracterização molecular objetivando estabelecer as relações filogenéticas existentes entre isolados de influenza aviária (H2N9) do Oriente Médio, outros isolados asiáticos e novas amostras dos Emirados Árabes Unidos coletadas de 2000 a 2003 foi realizado. Após a epidemia causada por H5N1, os autores consideraram importante a caracterização

de outras linhagens virais para investigação quanto ao seu potencial pandêmico. Para a inferência filogenética, oito segmentos de genes foram analisados. Das sete novas amostras estudadas, todas tiveram a adição de um aminoácido na hemaglutinina (HA) e seis continham uma mutação associada ao aumento da afinidade com os receptores de células humanas. Os genes que codificam para as glicoproteínas de superfície dos vírus e para a maioria das proteínas internas tiveram 90% de similaridade com uma linhagem de Hong Kong que causou mortes em crianças em 1997. O gene que codifica a neuraminidase teve quatro substituições, como aconteceu com os vírus pandêmicos de humanos. A sequência do fragmento M2 mostrou a troca de dois aminoácidos, o que está associado ao aumento de resistência à amantadina em cultivo celular. Devido às características supracitadas é sugerido pelos autores que H2N9 seja potencialmente pandêmico [1]. Ainda sobre influenza, citamos como exemplo uma nota prévia [16] que apresentou estudo com 24 sequências de influenza A (H1N2) isoladas de suínos, nos EUA, de 1999 a 2001, de propriedades com casos clínicos da doença. Os autores tinham como objetivo caracterizar oito segmentos do genoma de cada uma das amostras virais na tentativa de rastrear sua origem. A espécie suína possui em seu epitélio traqueal receptores para o vírus influenza com origem aviária e mamífera. Portanto, suínos são extremamente importantes no que diz respeito a recombinações entre diferentes cepas de influenza, quando sujeitos à co-infecção. Os resultados demonstraram que as 24 amostras estudadas são recombinações entre o H1N1 clássico e a influenza suína (H3N2), reforçando a necessidade de vigilância epidemiológica. Outro estudo [44] utilizou 156 genomas completos de influenza A (H3N2) de casos ocorridos em humanos de 1999 a 2004 no estado de Nova York nos EUA. Filogenias foram geradas para cada segmento genético do vírus, as quais revelaram que múltiplas recombinações ocorreram. O dado mais relevante encontrado foi que, no *cluster* que incluía amostras de H3N2, verificou-se que a hemaglutinina de um H3N2 coletado em 2000 foi passada por recombinação para todos os demais H3N2 amostrados de 2002 a 2003. Essa nova linhagem foi identificada como a causadora da epidemia 2002/2004. Além disso, constatou-se que diversas outras linhagens de influenza continuam co-circulando na mesma população.

As filogenias moleculares para caracterização da diversidade molecular têm sido descritas para uma grande variedade de microrganismos: Febre Hemorrágica Crimeia-Congo (CCHF) [55]; Vírus da Imunodeficiência Símia (SIV) [60]; Vírus da Hepatite B [45]; Vírus da Encefalomiocardite [56]; Vírus Símeo Espumoso [47]; Rotavírus Aviário grupo A [95]; Hantavírus [48]; Vírus da Anemia Aviária [97]; Vírus da Cinomose [12]; Vírus da Diarreia Viral Bovina (BVDV) [112]; Herpesvírus [84]; Vírus da Bronquite Infeciosa Aviária [34].

2. Dinâmica Populacional

A estrutura genética da população de patógenos humanos quase sempre reflete padrões conhecidos de migração humana. No caso de populações animais, parasitas transmitidos de forma direta, algumas vezes, fornecem informações substanciais a respeito de características temporais e espaciais do contato parasito-hospedeiro [82]. Estudos que buscam informações a respeito de estrutura populacional e história demográfica, entre outros, utilizam análises filogenéticas de parasitas para rastrear informações sobre seus hospedeiros.

2.1 Vírus da Imunodeficiência Felina (FIV)

FIV é um lentivírus que infecta felídeos no mundo inteiro. A sua alta taxa evolutiva contribui para uma ou mais mutações em cada nova partícula viral. O efeito disso é uma grande diversidade de linhagens virais.

Estudando leões (*Panthera Leo*) na África, Antunes *et al* [3], utilizaram marcadores moleculares e o gene *pol* do vírus da imunodeficiência felina (FIV - *Feline Immunodeficiency Virus*) para fazer a caracterização da variabilidade genética, traçar a história demográfica e evolutiva dos leões e avaliar aspectos da conservação dessa espécie animal. Com relação a FIV, as análises filogenéticas mostraram a existência de seis subtipos circulantes, contrariando a hipótese de que os leões africanos são provenientes de uma única população e adicionalmente refletindo um cenário histórico de refúgios/isolamentos e posterior expansão populacional. Ainda, o estudo demonstrou que marcadores virais muitas vezes concordam com marcadores genômicos quando se infere cenários evolutivos de uma espécie animal e podem ser muito informativos na relação tempo e espaço. Uma explicação bastante simplificada para avaliar estes cenários é que indivíduos

precisam de contato para reproduzir e transferir seu material genético. O contato direto ou indireto também é necessário para transmissão de vírus e no caso do FIV em populações silvestres, o meio de transmissão mais provável é através de brigas entre indivíduos. Se uma população está isolada em uma escala de tempo e espaço, os vírus desta população se diferenciam, formando novas linhagens virais. Em um outro estudo, por BIEK *et al* [7], a rápida evolução do FIV foi utilizada para revelar detalhes de estrutura populacional recente do puma (*Puma concolor*). Pumas são animais com alta prevalência de infecção por FIV, sem aparentemente desenvolverem a doença, sendo que a transmissão pode ser horizontal ou vertical, exigindo contato direto e causando infecções crônicas. Mais de trezentos animais de duas diferentes regiões foram estudados: 27,8% (98/352) deles estavam contaminados com o vírus, 7 foram encontrados com co-infecção dos diferentes subtipos de FIV, o que poderia levar à existência de recombinantes. As análises filogenéticas geradas revelaram a divisão das amostras em dois *clusters* principais e oito linhagens; os *clusters* correspondiam às duas regiões geográficas distintas estudadas. Em uma delas foi encontrada menor variabilidade, que deve ter ocorrido devido a uma redução na população de pumas em decorrência do aumento intensivo da caça desses animais nessa localidade, indicando que a dinâmica populacional do hospedeiro interfere na dinâmica da população viral.

3. Evolução Molecular

3.1 Retrovírus

Os retrovírus (família *Retroviridae*) consistem em um grupo diverso de vírus RNA e são importantes agentes de doenças para humanos (vírus da imunodeficiência humana, vírus linfotrópico da célula T humana) e animais (vírus da anemia infecciosa equina, vírus da imunodeficiência felina, vírus da leucose aviária, dentre outros). A marca diferenciada desses vírus é a sua forma de replicação, que consiste na sua transcrição reversa (RNA transcrito para DNA) e posterior integração no genoma do hospedeiro. As estimativas sobre existência dos retrovírus são bastante antigas chegando 100 milhões de anos, no caso dos Espumavírus [52]. Consegue-se estimar datas da existência desta família porque muitas sequências estão integradas no genoma de seus hospedeiros, como se fossem verdadeiros fósseis, denominados retrovírus endógenos [62].

Neste caso, a técnica molecular para datação consiste na comparação das sequências de nucleotídeos entre as duas extremidades longas e repetidas do provírus (LTRs - *Long Terminal Repeats*) que são idênticas no momento da integração. Muitas vezes é possível também inferir o local de origem de uma linhagem viral, além da data mais provável de seu surgimento.

A combinação de dados genômicos, ambientais e clínicos têm ajudado na caracterização da origem, transmissão e evolução, não só do HIV, mas de várias outras doenças infecciosas, como por exemplo Influenza A. Entretanto, vale a pena ressaltar que os resultados desses estudos dependem da amostragem, além de incertezas nas estimativas filogenéticas [35].

3.2. Vírus Influenza A

Os vírus estão em constante evolução e o vírus da influenza pode adquirir novidades evolutivas através de mecanismos como mutação pontual, recombinação e reassortimento. Este último mecanismo ocorre quando dois vírus diferentes infectam a mesma célula permitindo a criação de um novo agente, um verdadeiro mosaico. Entretanto, uma hipótese gerada por estudos ecológicos afirma que linhagens de influenza, que possuam como reservatórios aves silvestres aquáticas estão em estase evolucionária; ou seja, em baixas taxas evolutivas. Essa hipótese foi avaliada [15] estimando-se taxas de substituição de nucleotídeos em diversas amostras de influenza aviária. O autor chega à conclusão de que a hipótese acima não estava correta, pois as taxas de substituição de nucleotídeos encontradas eram extremamente altas, em torno de 10^{-3} substituições por sítio por ano. Os comprimentos dos ramos das árvores geradas confirmam essa informação, sugerindo que os vírus de aves silvestres evoluem mais rapidamente quando comparados aos de aves domésticas e que a diversidade genética encontrada também é muito grande dentro de cada subtipo e geralmente de origem recente. No entanto, os ciclos evolutivos de influenza em aves silvestres e domésticas não podem ser analisados em separado, pois a movimentação dos vírus entre as diferentes espécies de aves é constante. Outro bom exemplo de aplicações de análises filogenéticas para estudos de evolução e recombinações é um trabalho [98] que investigou a origem evolucionária de amostras da influenza A (H1N1), recente causadora de pandemia. O objetivo dos autores era estimar o tempo que foi necessário para o aparecimento dessa nova linhagem de influenza, de que forma ela surgiu e

quanto tempo levou para se tornar epidêmica. Diversas sequências de influenza de diferentes espécies animais foram utilizadas no estudo, as quais foram amostradas do *NCBI Influenza Virus Resource*. Os resultados demonstraram que esse novo vírus é proveniente de cepas comuns à espécie suína. Os genes HA, NP e NS são provenientes de recombinação com influenza suína clássica, PB2 e PA derivados de influenza aviária, PB1 vindo do H3N2 humano, que deram origem a uma linhagem que circula há anos na América do Norte. Em contraste, os genes NA e M da nova gripe são provenientes de uma linhagem de H1N1 de aves da Eurásia. Além disso, os dados obtidos mostram que as primeiras transmissões a humanos ocorreram vários meses antes do início da epidemia e destacam a necessidade de uma vigilância sistemática de influenza em suínos, considerando que essa espécie pode servir para novas recombinações que poderiam gerar novos organismos com potencial pandêmico. Diversos bons exemplos de estudos filogenéticos que auxiliam na compreensão da evolução dos patógenos podem ser encontrados na literatura [44,63,69,113].

VI. CONCLUSÃO

A sequência de DNA não é só um registro de relações filogenéticas e tempos de divergência, mas também guarda marcas de como o processo evolucionário deu forma a sua história e também o tempo

decorrido no processo evolutivo das populações [72]. Cientistas estão expandindo o uso de filogenias para um número muito grande de finalidades e estão traçando a diversificação de linhagens em diferentes classes, famílias, gêneros, espécies e dentro de um único indivíduo. Na veterinária, a utilização desta ferramenta auxilia na elaboração de programas de controle e erradicação de doenças, sendo importante para rastrear os locais ou espécies animais de onde comumente surgem novos microrganismos, como vimos nas epidemias de influenza. Os estudos de dinâmica populacional, possibilitam relacionar a variabilidade genética do patógeno com a variabilidade genética do hospedeiro. A filogenia para caracterização molecular, estabelecendo o conhecimento das linhagens circulantes em determinados locais, fornece subsídios para elaborar estratégias de imunização condizentes com as necessidades reais.

As análises de sequências genômicas por meio de filogenias moleculares têm demonstrado um amplo espectro de aplicações. No entanto, é importante ressaltar que para os diferentes dados disponíveis e diferentes finalidades das filogenias, os métodos de reconstrução e modelos evolutivos devem ser escolhidos de forma criteriosa. A escolha da melhor forma de fazer uma análise filogenética está relacionada, principalmente, com os dados de que se dispõe; além disso, é necessário um alinhamento confiável e escolha de modelo evolutivo apropriado.

REFERENCES

- 1 Aamir U.B., Wernery U., Ilyushina N. & Webster R.G. 2007. Characterization of avian H9N2 influenza viruses from United Arab Emirates 2000 to 2003. *Virology*. 361(1): 45-55.
- 2 Abascal F., Zardoya R. & Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21(9): 2104-2105.
- 3 Antunes A., Troyer J.L., Roelke M.E., Pecon-Slattery J., Packer C., Winterbach C., Winterbach H., Hemson G., Frank L., Stander P., Siefert L., Driciru M., Funston P. J., Alexander K.A., Prager K.C., Mills G., Wildt D., Bush M., O'Brien S.J. & Johnson W.E. 2008. The evolutionary dynamics of the lion (*Panthera leo*) revealed by host and viral population genomics. *PLOS Genetics*. 4(11): e1000251.
- 4 Batista W.C., Kashima S., Marques A.C. & Figueiredo L.T.M. 2001. Phylogenetic analysis of Brazilian Flavivirus using nucleotide sequences of parts of NS5 gene and 3' non-coding regions. *Virus Research*. 75(1): 35-42.
- 5 Bayes T. 1763. An essay towards solving in the doctrine of chances. *Philosophical Transactions of the Royal Society London*. 53(1): 370-418.
- 6 Bhullar K., Waglechner N., Pawlowski A., Koteva K. & Banks E.D. 2012. Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome. *PLoS ONE*. 7(4): e34953.
- 7 Biek R., Drummond A.J. & Poss M. 2006. A virus reveals population structure and recent demographic history of its carnivore host. *Science*. 311(5760): 538-541.
- 8 Brandão P.E., Gregori F., Richtzenhain L.J., Rosales C.A.R., Villareal L.Y.B. & Jerez J.A. 2006. Molecular diversity of Brazilian strains of bovine coronavirus (BCoV) reveals a deletion within the hypervariable region of the S1 subunit of the spike glycoprotein also found in human coronavirus OC43. *Archives of Virology*. 151(9): 1735-1748.

- 9 Brandão P.E., Scheffer K., Villarreal L.Y., Achkar S., Oliveira R.N., Fahl W.O., Castilho J.G., Kotait I. & Richtzenhain L.J. 2008. A coronavirus detected in the vampire bat *Desmodus rotundus*. *Brazilian Journal of Infectious Diseases*. 12(6): 466-468.
- 10 Brandão P.E., Villareal L.Y.B., Gregori F., Souza S.L.P., Lopes M.A.E., Gomes C.R., Sforsin A.J., Sanches A.A., Rosales C.A.R Richtzenhain L.J., Ferreira A.J.P. & Jerez J.A. 2007. On the etiology of an outbreak of winter dysentery in dairy cows in Brazil. *Pesquisa Veterinária Brasileira*. 27(10): 398-402.
- 11 Brocchieri L. 2001. Phylogenetic Inferences from Molecular Sequences: Review and Critique. *Theoretical Population Biology*. 59(1): 27-40.
- 12 Budaszewski R.F., Pinto L.D., Weber M.N., Caldart E.T., Alves C.D.B.T., Martellac V., Ikutad N., Lunged V.R. & Canal C.W. 2014. Genotyping of canine distemper virus strains circulating in Brazil from 2008 to 2012. *Virus Research*. 180(1): 76-83. DOI: 10.1016/j.virusres.2013.12.024.
- 13 Carnieli Jr. P., Brandão P.E., Carrieri M.L., Castilho J.G., Macedo C.I., Machado L.M., Rangel N., de Carvalho R.C., de Carvalho V.A., Montebello L., Wada M. & Kotait I. 2006. Molecular epidemiology of rabies virus strains isolated from wild canids in Northeastern Brazil. *Virus Research*. 120(1-2):113-120.
- 14 Cavalli-Sforza L.L. & Edwards A.W.F. 1967. Phylogenetic Analysis Models and Estimation Procedures. *American Journal of Human Genetics*. 19(3): 233-257.
- 15 Chen R. & Holmes E.C. 2006. Avian Influenza Virus exhibits rapid evolutionary dynamics. *Molecular Biology and Evolution*. 23(12): 2336-2341.
- 16 Choi Y.K., Goyal S.M., Farnham M.W. & Joo H.S. 2002. Phylogenetic analysis of H1N2 isolates of Influenza A Virus from pigs in the United States. *Virus Research*. 87(2): 173-179.
- 17 Cisterna D., Bonaventura R., Caillou S., Pozo O., Andreau M.L., Fontana L.D., Echegoyen C., de Mattos C., de Mattos C., Russo S., Novaro L., Elbergerh D. & Freire M.C. 2005. Antigenic and molecular characterization of rabies virus in Argentina. *Virus Research*. 109(2): 139-147.
- 18 Davies J. & Davies D. 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*. 74(3): 417-433.
- 19 De Robertis E.M.F. & Hib J. 2006. *Bases da biologia celular e molecular*. 4.ed. Rio de Janeiro: Guanabara Koogan, 389p.
- 20 Dorman J.S. 2000. Molecular epidemiology: The impact of molecular biology in epidemiology research. *Revista Medica de Chile*. 128(11): 1261-1268.
- 21 Edwards A.W.F. 1972. *Likelihood. An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference*. New York: Cambridge University Press, 236p.
- 22 Efron B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*. 7(1): 1-26.
- 23 Étienne J., Millot F. & Cerqueira A.J. 2003. *Bioquímica genética e biologia molecular*. 6.ed. São Paulo: Livraria & Editora, 504p.
- 24 Ewald P.W. 2004. Evolution of virulence. *Infectious Disease Clinics of North America*. 18(1): 1-15.
- 25 Ewing B. & Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research*. 8(1): 186-194.
- 26 Ewing B., Hillier L. & Wendl M.C. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*. 8(1): 175-185.
- 27 Feitosa A.L.V.L., da Silva Teixeira M.F., Pinheiro R.R., da Cunha R.M.S., Lima J.P.M.S., Andrioli A. & Pinheiro D.C.S.N. 2010. Phylogenetic analysis of small ruminant lentiviruses from Northern Brazil. *Small Ruminant Research*. 94(1): 205-209.
- 28 Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*. 27(4): 401-410.
- 29 Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. 17(6): 368-376.
- 30 Felsenstein J. 1985. Confidence limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 39(4): 783-791.
- 31 Felsenstein J. 2004. *Inferring Phylogenies*. 2nd edn. Sunderland: Sinauer Associates Inc., 664p.
- 32 Ferreira H.B., Passaglia L.M.P. & Zaha A. 2003. *Biologia molecular básica*. 3.ed. Porto Alegre: Mercado Aberto, 421p.

- 33 Forattini O.P. 2004.** *Conceitos básicos de epidemiologia molecular*. São Paulo: Editora da Universidade de São Paulo, 98p.
- 34 Fraga A.P., Balestrin E., Ikuta N., Fonseca A.S.K., Spilki F.R., Canal C.W. & Lunge V.R. 2013.** Emergence of a New Genotype of Avian Infectious Bronchitis Virus in Brazil. *Avian Diseases*. 57(2): 225-232.
- 35 Frost S.D.W., Pybus O.G., Gog J.R., Viboud C., Bonhoeffer S. & Bedford T. 2015.** Eight challenges in phylodynamic inference. *Epidemics*. 10: 88-92.
- 36 Garcia M. 2007.** Uma filogenia mitocondrial de metazoários. 212f. Petrópolis, RJ. Dissertação (Mestrado em Modelagem Computacional com Ênfase em Bioinformática) - Programa de Pós-graduação em Modelagem Computacional, Universidade Federal do Rio de Janeiro.
- 37 Gilks W.R., Richardson S. & Spiegelhalter D.J. 1996.** *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall, 512p.
- 38 Goldman N. & Whelan S. 2000.** Statistical tests of γ -distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Molecular Biology and Evolution*. 17(1): 974-978.
- 39 Grego E., Profiti M., Giammarioli M., Giannino L., Rutili D., Woodall C., & Rosati S. 2002.** Genetic heterogeneity of small ruminant lentiviruses involves immunodominant epitope of capsid antigen and affects sensitivity of single-strain-based immunoassay. *Clinical and diagnostic laboratory immunology*. 9(4): 828-832.
- 40 Hasegawa M., Kishino H. & Yano T. 1985.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*. 22(2): 160-174.
- 41 Hillis D.M. & Bull J.J. 1993.** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*. 42(2): 182-192.
- 42 Holder M. & Lewis P.O. 2003.** Phylogeny estimation: traditional and Bayesian approaches. *Nature*. 4(4): 275-284.
- 43 Holmes E.C. 1998.** Molecular epidemiology and evolution of emerging infectious diseases. *British Medical Bulletin*. 54(3): 533-543.
- 44 Holmes E.C., Ghedin E., Miller N., Taylor J., Bao Y., George K.S., Grenfell B.T., Salzberg S.L., Fraser C.M., Lipman D.J. & Taubenberger J.K. 2005.** Whole-genome analysis of Human Influenza A Virus reveals multiple persistent lineages and reassortment among recent H3N2 Virus. *PLoS*. 3(9): 1579-1589.
- 45 Hu X., Javadian A., Gagneux P. & Robertson B.H. 2001.** Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. *Virus Research*. 79(1): 103-108.
- 46 Huelsenbeck J.P., Ronquist F., Nielsen R. & Bollback J.P. 2001.** Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*. 294(5550): 2310-2314.
- 47 Hussain A.I., Shanmugam V., Bhullar V.B., Beer B.E., Vallet D., Gautier-Hion A., Wolfe N.D., Karesh W.B., Kilbourn A.M., Tooze Z., Heneine W. & Switzer W.E. 2003.** Screening of simian foamy virus infection by using a combined antigen western blot assay: evidence for a wide distribution among old world primates and identification of four new divergent viruses. *Virology*. 309(2): 248-257.
- 48 Jakab F., Horváth G., Ferenczi E., Sebók J., Varcza Z. & Szúcs G. 2007.** Detection of Dobrava hantaviruses in *Apodemus agrarius* mice in the Transdanubian region of Hungary. *Virus Research*. 128(1-2): 149-152.
- 49 Jukes T.H. & Cantor C.R. 1969.** Evolution of Protein Molecules. In: Munro H.N. (Ed). *Mammalian Protein Metabolism*. 3rd edn. New York: Academic Press, pp.21-132.
- 50 Junqueira L.C.U. & Carneiro J. 2005.** *Biologia celular e molecular*. 8.ed. Rio de Janeiro: Guanabara Koogan, 332p.
- 51 Karr B. M., Chebloune Y., Leung K. & Narayan O. 1996.** Genetic characterization of two phenotypically distinct North American ovine lentiviruses and their possible origin from caprine arthritis-encephalitis virus. *Virology*. 225(1): 1-10.
- 52 Katzourakis A., Aiewsakun P., Jia H., Wolfe N.D., LeBreton M. & Yoder A.D. 2014.** Discovery of prosimian and afrotherian foamy viruses and potential cross species transmissions amidst stable and ancient mammalian co-evolution. *Retrovirology*. 11(1): 1.
- 53 Khuner M.K. & Felsenstein J. 1994.** A Simulation Comparison of Phylogeny Algorithms Under Equal and Unequal Evolutionary Rates. *Molecular Biology and Evolution*. 11(3): 459-468.
- 54 Kimura M. 1980.** A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 16(1): 111-120.

- 55 Kinsella E., Martin S.G., Grolla A., Czub M., Feldmann H. & Flicka R. 2004. Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology*. 321(1): 23-28.
- 56 Knowles N.J., Dickinson N.D., Wilsden G., Carra E., Brocchi E. & De Simone F. 1998. Molecular analysis of encephalomyocarditis viruses isolated from pigs and rodents in Italy. *Virus Research*. 57(X): 53-62.
- 57 Leroux C., Chastang J., Greenland T. & Mornex J.F. 1997. Genomic heterogeneity of small ruminant lentiviruses: existence of heterogeneous populations in sheep and of the same lentiviral genotypes in sheep and goats. *Archives of virology*. 142(6): 1125-1137.
- 58 Leser W. 1985. *Elementos em epidemiologia geral*. Rio de Janeiro: Atheneu, 178p.
- 59 Levin B.R., Lipsitch M. & Bonhoeffer S. 1999. Population Biology, Evolution, and Infectious Disease: Convergence and Synthesis. *Science*. 283(5403): 806-809.
- 60 Liegeois F., Courgnaud V., Switzer W.M., Murphy H.W., Loul S., Aghokeng A., Pourrut X., Mpoudi-Ngole E., Delaporte E. & Peeters M. 2006. Molecular characterization of a novel simian immunodeficiency virus lineage (SIVtal) from northern talapoin (*Miopithecus ogouensis*). *Virology*. 349(1): 55-65.
- 61 Mata H., Fontana C.S., Maurício G.N., Bornschein M.R., de Vasconcelos M.F. & Bonatto S.L. 2009. Molecular phylogeny and biogeography of the eastern Tapaculos (Aves: Rhinocryptidae: *Scytalopus*, *Eleoscytalopus*): Cryptic diversification in Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*. 53(2): 450-462. DOI:10.1016/j.ympev.2009.07.017.
- 62 Mata H., Gongora J., Eizirik E., Alves B.M., Soares M.A. & Ravazzolo A.P. 2015. Identification and characterization of diverse groups of endogenous retroviruses in felids. *Retrovirology*. Mar 15;12:26. DOI: 10.1186/s12977-015-0152-x.
- 63 Mata H., Gongora J. & Ravazzolo A.P. 2013. Molecular Characterization of SINEs Integrated in Endogenous Retrovirus Sequences from *Leopardus geoff royi* and *Puma concolor*. *Acta Scientiae Veterinariae*. 41(1): 1133.
- 64 Matioli S.R. 2001. *Biologia molecular e evolução*. Ribeirão Preto: Holos, 202p.
- 65 Maruyama S.R., Castro-Jorge L.A., Ribeiro J.M.C., Gardinassi L.G., Garcia G.R., Brandão L.G., Rodrigues A.R., Okada M.I., Abrão E.P., Ferreira B.R., Fonseca B.A.L. & Miranda-Santos I.K.F. 2014. Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in *Rhipicephalus microplus* ticks from Brazil. *Memórias do Instituto Oswaldo Cruz*. 109(1): 38-50.
- 66 Maurício G.N., Mata H., Bornschein M.R., Cadena C.D., Alvarenga H. & Bonatto S.L. 2008. Hidden generic diversity in Neotropical birds: Molecular and anatomical data support a new genus for the “*Scytalopus*” *indigoticus* species-group (Aves: Rhinocryptidae). *Molecular Phylogenetics and Evolution*. 49(1): 125-135.
- 67 Metzker M.L., Mindell D.P., Liu X-M., Ptak R.G., Gibbs R.A. & Hillis D.M. 2002. Molecular evidence of HIV-1 transmission in a criminal case. *PNAS*. 99(22): 14292-14297.
- 68 Nei M. 1996. Phylogenetic Analysis in Molecular Evolutionary Genetics. *Annual Review of Genetics*. 30(1): 371-403.
- 69 O'Brien S.J., Troyer J.L., Roelke M., Markerc L. & Pecon-Slattery J. 2006. Plagues and adaptation: lessons from the felidae models for SARS and AIDS. *Biological Conservation*. 131(2): 255-267.
- 70 Ota R., Waddell P.J., Hasegawa M., Shimodaira H. & Kishino H. 2000. Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. *Molecular Biology and Evolution*. 17(5): 798-803.
- 71 Ou C.Y., Ciesielski C.A., Myers G., Bandea C.I., Luo C.-C., Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L., Economou A.N., Witte J.J., Furman L.J., Satten G.A., Maclnnes K.A., Curran J.W. & Jaffe H.W. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science*. 256(5060): 1165-1171.
- 72 Page R.D.M. & Holmes E.C. 1998. *Molecular evolution: a Phylogenetic Approach*. Oxford: Blackwell Science Ltd., 352p.
- 73 Pennings P.S. 2012. Standing Genetic Variation and the Evolution of Drug Resistance in HIV. *PLoS Computacional Biology*. 8(6): e1002527.
- 74 Pinto L.D., Barros I.N., Budaszewski R.F., Weber M.N., Mata H., Antunes J.R., Boabaid F.M., Wouters A.T.B., Driemeier D., Brandão P.E. & Canal C.W. 2014. Characterization of pantropic canine coronavirus from Brazil. *The Veterinary Journal*. 202(3): 659-662. DOI:10.1016/j.tvjl.2014.09.006.
- 75 Pisoni G., Bertoni G., Boettcher P., Ponti W.A. & Moroni P. 2006. Phylogenetic analysis of the *gag* region encoding the matrix protein of small ruminant lentiviruses: comparative analysis and molecular epidemiological applications. *Virus Research*. 116(1): 159-167.

- 76 **Pisoni G., Bertoni G., Puricelli M., Maccalli M. & Moroni P. 2007.** Demonstration of coinfection with and recombination by caprine arthritis-encephalitis virus and maedi-visna virus in naturally infected goats. *Journal of Virology*. 81(10): 4948-4955.
- 77 **Posada D. 2008.** jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*. 25(7): 1253-1256.
- 78 **Posada D. 2009.** Selecting models of evolution. In: Salemi M. (Ed). *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*. 2nd edn. Cambridge: Cambridge University Press, pp.256-282.
- 79 **Posada D. & Crandall K.A. 2001.** Selecting the best-fit model of nucleotide substitution. *Systematic Biology*. 50(4): 580-601.
- 80 **Rannala B. & Yang Z. 1996.** Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference. *Journal of Molecular Evolution*. 43(3): 304-311.
- 81 **Ravazzolo A.P., Reischak D., Peterhans E. & Zanoni R. 2001.** Phylogenetic analysis of small ruminant lentiviruses from Southern Brazil. *Virus Research*. 79(1): 117-123. DOI:10.1016/S0168-1702(01)00339-2.
- 82 **Real L.A. & Biek R. 2007.** Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *Journal of the Royal Society*. 4(16): 935-948.
- 83 **Riley L.W. 2004.** *Molecular epidemiology of infectious diseases: principles and practices*. Washington: ASM Press, 348p.
- 84 **Rodenbusch C.R., Baptistotte C., Werneck M.R., Pires T.T., Melo M.T.D., de Ataíde M.W., dos Reis K.D.H.L., Testa P., Alievi M.M. & Canal C.W. 2014.** Fibropapillomatosis in green turtles *Chelonia mydas* in Brazil: characteristics of tumors and virus. *Diseases of Aquatic Organisms*. 111(3): 207-217. DOI:10.3354/dao02782.
- 85 **Rokas A. & Carrol S.B. 2005.** More Genes or More Taxa? The Relative Contribution of Gene Number and Taxon Number to Phylogenetic Accuracy. *Molecular Biology and Evolution*. 22(5): 1337-1344.
- 86 **Ronquist F. & Deans A.R. 2010.** Bayesian phylogenetics and its influence on insect systematics. *Annual Review of Entomology*. 55(1): 189-206.
- 87 **Ronquist F., Huelsenbeck J.P. & Britton T. 2004.** Bayesian supertrees. In: Bininda-Emonds O.R.P. (Ed). *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life*. Amsterdam: Kluwer, pp.193-224.
- 88 **Ronquist F., Teslenko M., Mark P.V.D., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012.** MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*. 61(3): 539-542.
- 89 **Rost B. 1999.** Twilight zone of protein sequence alignments. *Protein Engineering*. 12(2): 85-94.
- 90 **Russel B. 1946.** *A History of Western Philosophy*. London: Routledge, 895p.
- 91 **Saitou N. & Imanishi T. 1989.** Relative Efficiencies of Fitch-Margoliash, Maximum-Parsimony, Maximum-Likelihood, Minimum-Evolution, and Neighbor-Joining Methods of Phylogenetic Tree Construction in Obtaining the Correct Tree. *Molecular Biology and Evolution*. 4(4): 406-425.
- 92 **Saitou N. & Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 4(4): 406-425.
- 93 **Saltarelli M., Querat G., Konings D.A., Vigne R. & Clements J.E. 1990.** Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. *Virology*. 179(1): 347-364.
- 94 **Schneider H. 2003.** *Métodos de Análise filogenética: um guia prático*. Ribeirão Preto: Holos Editora, 162p.
- 95 **Schumann T., Hotzel H., Otto P. & Johne R. 2009.** Evidences of interespecies transmission and reassortment among avian group A rotaviruses. *Virology*. 386(2): 334-343.
- 96 **Shah C., Böni J., Huder J.B., Vogt H.R., Mühlherr J., Zanoni R. & Schüpbach J. 2004.** Phylogenetic analysis and reclassification of caprine and ovine lentiviruses based on 104 new isolates: evidence for regular sheep-to-goat transmission and worldwide propagation through livestock trade. *Virology*. 319(1): 12-26.
- 97 **Simionatto S., Lima-Rosa C.A.V., Binneck E., Ravazzolo A.P. & Canal C.W. 2006.** Characterization and phylogenetic analysis of Brazilian chicken anaemia virus. *Virus Genes*. 33(1): 5-10. DOI: 10.1007/s11262-005-0033-9.
- 98 **Smith G.J.D., Vijaykrishna D., Bahl J., Lycett S.J., Worobey M., Pybus O.G., Ma S.M., Cheung C.L., Raghwani J., Bhatt S., Peiris J.S.M. & Guan Y., Rambaut A. 2009.** Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. 459(7250): 1122-1126.
- 99 **Sonigo P., Alizon M., Staskus K., Klatzmann D., Cole S., Danos O., Retzel E., Tiollais P., Haase A. & Wain-Hobson S. 1985.** Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell*. 42(1): 369-382.

- 100 Sounis E. 1985.** *Epidemiologia geral*. São Paulo: Atheneu, 178p.
- 101 Steel M. & Penny D. 2000.** Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution*. 17(6): 839-850.
- 102 Studier J.A. & Keppler K.J. 1988.** A note on the neighbor-joining algorithm of Saitou and Nei. *Molecular Biology and Evolution*. 5(6): 729-731.
- 103 Tajima F. & Nei M. 1984.** Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*. 1(3): 269-285.
- 104 Takahashi K. & Nei M. 2000.** Efficiencies of Fast Algorithms of Phylogenetic Inference Under the Criteria of Maximum Parsimony, Minimum Evolution, and Maximum Likelihood When a Large Number of Sequences Are Used. *Molecular Biology and Evolution*. 17(3): 1251-1258.
- 105 Tamura K. 1992.** Estimation of the number of nucleotide substitutions when there are strong transition-transversion and GC-content biases. *Molecular Biology and Evolution*. 9(4): 678-687.
- 106 Tamura K. & Nei M. 1993.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*. 10(3): 512-526.
- 107 Thrusfield M. 2004.** *Epidemiologia Veterinária*. 2.ed. São Paulo: Roca, 556p.
- 108 Turner P.C. & Motta P.A. 2004.** *Biologia molecular*. Rio de Janeiro: Guanabara Koogan, 287p.
- 109 Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. 2001.** Genetic evidence for Near-Eastern origins of European cattle. *Nature*. 410(6832): 1091-1110.
- 110 Vandamme A.M. 2009.** Basic Concepts of Molecular Evolution. In: Salemi M. (Ed). *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*. 2nd edn. Cambridge: Cambridge University Press, pp.1-23.
- 111 Villareal L.Y.B., Brandão P.E., Chacón J.L., Saldenberga A.B.S., Assayag A.S., Jones R.C. & Ferreira A.J.P. 2007.** Molecular characterization of infectious bronchitis virus strains isolated from the enteric contents of brazilian laying hens and broilers. *Avian Diseases*. 51(4): 974-978.
- 112 Weber M.N., Silveira S., Machado G., Groff F.H.S., Mósen A.C.S., Budaszewski R.F., Dupont P.M., Corbellini L.G. & Canal C.W. 2014.** High frequency of bovine viral diarrhoea virus type 2 in Southern Brazil. *Virus Research*. 191: 197-124.
- 113 Weber M.N., Streck A.F., Silveira S., Mósen A.C.S., da Silva M.S. & Canal C.W. 2015.** Homologous recombination in pestiviruses: Identification of three putative novel events between different subtypes/genogroups. *Infection, Genetics and Evolution*. 30: 219-224. DOI:10.1016/j.meegid.2014.12.032.
- 114 Whelan S., Liò P. & Goldman N. 2001.** Molecular phylogenetics: state-of-the art methods for looking into the past. *Trends in Genetics*. 17(5): 262-272.
- 115 World Health Organization (WHO). 2016.** Disponível em: <<http://www.who.int/topics/en/>>. [Accessed April 2016.]
- 116 Xia X. & Xie Z. 2001.** DAMBE: data analysis in molecular biology and evolution. *Journal of Heredity*. 92(4): 371-373.
- 117 Yang Z. 1994.** Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution*. 39(1): 105-111.
- 118 Zharkikh A. & Li W.-H. 1992.** Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Journal of Molecular Evolution*. 9(6): 1119-1147.